

RESEARCH ARTICLE

Toward Label-Efficient Neural Network Training: Diversity-Based Sampling in Semi-Supervised Active Learning

FELIX BUCHERT¹, NASSIR NAVAB^{1,2}, (Fellow, IEEE), AND SEONG TAE KIM^{1,3}, (Member, IEEE)

¹Chair for Computer Aided Medical Procedures, Technical University of Munich, 85748 München, Germany

²Chair for Computer Aided Medical Procedures, Johns Hopkins University, Baltimore, MD 21218, USA

³Department of Computer Science and Engineering, Kyung Hee University, Yongin-si, Gyeonggi-do 17104, South Korea

Corresponding author: Seong Tae Kim (st.kim@khu.ac.kr)

This work was supported in part by the Kyung Hee University in 2021 under Grant KHU-20210732; in part by the National Research Foundation of Korea (NRF) Grant through the Korea Government (MSIT) under Grant 2021R1G1A1094990; and in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant through the Korea Government (MSIT), Artificial Intelligence Innovation Hub, under Grant 2021-0-02068.

ABSTRACT Collecting large-labeled data is an expensive and challenging issue for training deep neural networks. To address this issue, active learning is recently studied where the active learner provides informative samples for labeling. Diversity-based sampling algorithms are commonly used for representation-based active learning. In this paper, a new diversity-based sampling is introduced for semi-supervised active learning. To select more informative data at the initial stage, we devise a diversity-based initial dataset selection method by using self-supervised representation. We further propose a new active learning query strategy, which exploits both consistency and diversity. Comparative experiments show that the proposed method can outperform other active learning approaches on two public datasets.

INDEX TERMS Active learning, semi-supervised learning, data-efficient deep learning.

I. INTRODUCTION


Collecting large-labeled datasets is important for the success of modern deep neural networks. However, labeling from humans is generally expensive. In some applications where the labeling should be conducted by domain experts, building a large-labeled dataset is extremely difficult, which limits the development of deep learning. *Active learning* is a task that tries to address this issue where the active learning method suggests informative data for labeling to annotators under the limited budget of annotation.

The core component of active learning algorithm is the query strategy for quantifying the usefulness of unlabeled samples for selecting more useful samples for labeling at active learning cycles. Query strategies are used for representation-based active learning where the methods select data that encode more diverse representational

information. The *k-means++ initialization* [2], [4] and the *k-Center-Greedy algorithms* [33], [45] have been widely used for annotating a diverse set of samples. However, recent active learning studies have been explored on supervised learning setting where the unlabeled data is not used during training of the model.

Recently, semi-supervised learning studies are widely explored, which utilize both unlabeled data with labeled data for the training of the model. By designing methods to use unlabeled data during training, these studies improve the model, which can reduce the need for labeling for the success of deep learning. It is a natural choice to utilize both ideas of semi-supervised learning and active learning for the success of deep learning under the limited budget for the data labeling.

In this study, we introduce a new semi-supervised active learning method which suggests the data for labeling where the model is trained with semi-supervised learning conditions. Compared with previous active learning studies,

The associate editor coordinating the review of this manuscript and approving it for publication was Michele Magno .

the information gained from the labeling is changed under semi-supervised learning and it is an important task to predict the information quantity of data for successful training under the limited annotation budget. For this purpose, the proposed method consists of a diversity-based initial dataset selection based on self-supervised representation and a query strategy which uses both consistency and diversity for semi-supervised active learning. The proposed method combines consistency and diversity by using a consistency-based embedding scheme. Our contributions are summarized as follows:

- 1) A new initial dataset selection method based on the diversity of self-supervised representation is proposed for better starting active learning. The proposed approach yields significantly more informative initial datasets than traditional random selection methods.
- 2) A new active learning query strategy is designed for semi-supervised active learning. Central to the method is an embedding space that exploits both a sample's representational information and the consistency of model predictions.
- 3) Comparative experiments have been conducted to verify the effectiveness of the proposed method. Our method yields significant performance improvements on two public datasets.

A preliminary version of the paper was presented in [9]. By building upon [9], we further introduce the analyses to understand the characteristics of samples selected by the proposed method. To understand the samples from initial dataset selection approaches, class distribution imbalance and sample diversity are measured in random selection, k-means++ initialization step, and k-Center greedy algorithm (please see Figure 4). We further examine the distribution of sample embeddings which are calculated by self-supervised representation learning. TSNE approach [41] is used to visualize sample embeddings (please see Figure 5). To analyze the samples in the active learning query strategies, consistency and sample diversity are measured over different active learning approaches (please see Figure 6, Figure 7). Also, TSNE visualization is introduced to show the limitation of a purely consistency-based query strategy (please see Figure 2). These comprehensive analyses will improve the understanding of diversity-based sampling for semi-supervised active learning.

II. RELATED WORK

In the following, we discuss related work in the fields of semi-supervised learning, self-supervised learning and active learning with a particular focus on diversity-based sampling. As our work exclusively focuses on algorithms based on image classification using deep neural networks, we limit the discussion of related work to corresponding lines of research. For a broader introduction into these fields, we refer to the comprehensive survey of active learning [35] and introductory material on semi-supervised learning [10].

A. ACTIVE LEARNING

Recently studies on deep batch active learning follow one of two streams of approaches: Uncertainty-based methods and representation-based methods. Uncertainty-based active learning algorithms [14], [21], [30], [32], [40], [43] select data on which the current model's predictions are not confident. Entropy-based sampling [21] uses Shannon's entropy [36], an information-theoretic measure of uncertainty, to select samples for labeling, while margin-based algorithms [30], [32], [40] aim at selecting samples close to the current classifier's decision boundary. Monte Carlo (MC) dropout [14] is a more advanced approach for estimating model uncertainty.

In the representation-based active learning, a diverse batch of data which encode diverse representational information is selected as a representative of the entire dataset. Coreset [33] formulates representation-based sampling as core-set selection in a suitable embedding space and uses the k-Center-Greedy algorithm for sample selection.

In addition to active learning methods that solely rely on one of the previously discussed approaches, there are approaches that combine both uncertainty and diversity. *BADGE* [4] embeds unlabeled data into gradient embedding space. The representation of the data in the gradient embedding space is computed by multiplying the uncertainty by the feature. The k-means++ clustering initialization is used for selecting the samples on the gradient embedding space in [4].

B. SEMI-SUPERVISED LEARNING

Recent studies in semi-supervised learning have explored pseudo-labeling [25] and consistency regularization [31]. Mean Teacher [39] uses consistency regularization based on a teacher model. They utilize an exponential moving average of model weights over previous training iterations. *MixMatch* [7] employs the augmentation strategy *MixUp* [46] and exploit both pseudo-labeling and consistency regularization. Recently, *FixMatch* [37] combines pseudo-labeling and consistency regularization in a simple way. Using both weak and strong augmentation strategies to introduce a form of consistency regularization, *FixMatch* achieves state-of-the-art results across a variety of standard semi-supervised learning benchmarks.

C. SELF-SUPERVISED REPRESENTATION

Self-supervised learning is used to learn low-dimensional representations of data samples that extract useful information from data [5], [11], [12], [16], [28]. To utilize large-unlabeled data during training, the studies propose various pretext tasks. *RotNet* [16] predicts rotations of image by 0, 90, 180, and 270 degrees. It has proven to yield powerful image representations, which have been successfully used in downstream tasks such as image classification. *Context encoders* [28] learn semantic image representations by predicting large masked areas of an input image. *SimCLR* [11] leverages augmentations for self-supervised representation

learning. *BYOL* [18] relies on an online and a target network, which interact with each other in order to learn meaningful semantic representations.

Ideally, the self-supervised representations encode discriminative features, which makes them be useful in downstream tasks [5]. In the context of supervised learning, deep convolutional neural networks have shown to learn semantically meaningful image representations from large-labeled datasets on image classification tasks. The learned image representations generalize well and can be transferred to other vision tasks such as semantic segmentation [17] or image captioning [22]. Consequently, there has been an increased interest in learning good high-level representations of images in an unsupervised manner.

D. SEMI-SUPERVISED ACTIVE LEARNING

Semi-supervised learning and active learning are studies for minimizing the labeling effort required for neural network training. Therefore, research exploring the integration of these two approaches seems natural and promising. Semi-supervised active learning has been previously explored for time-series data, graphs, and natural language processing [8], [19], [20], [26], [27]. In image classification, proposed semi-supervised active learning algorithms have mostly been based on straightforward combinations of existing algorithms in both fields [12], [29], [33], [38]. However, only little research has been conducted into exploring active learning query strategies specifically constructed for the semi-supervised learning setting. Gao et al. [15] propose a consistency-based selection criterion, i.e. a criterion based on concepts used in semi-supervised learning, and report significant improvements over mere combinations of active and semi-supervised learning algorithms.

Following this line of research, in this study, we propose a new query strategy for the semi-supervised active learning, which utilizes the diversity-based sampling as well as a carefully constructed embedding space. Our method can automatically balance consistency and diversity based on consistency-based embedding. Moreover, we introduce a diversity-based initial dataset selection, which improves the performance of the model in the follow-up active learning steps.

III. DIVERSITY-BASED SAMPLING ALGORITHMS

Diversity-based sampling algorithms have been widely used in active learning. The k-means++ initialization [4] and the k-Center-Greedy algorithm [33], [45] are representative approaches. Let \mathcal{D} denote the set of labeled data. The diversity-based sampling iteratively select data based on their distance to the closest selected data. For the i -th sample, the distance is calculated as

$$d_i = \min_{j \in \mathcal{D}} \|f_i - f_j\|_2, \quad (1)$$

where d_i represents the L2-distance between the features of data f_i , and the features of the closest data in \mathcal{D} . The data is

selected as

$$p(f = f_i) = \frac{d_i^{\frac{1}{T}}}{\sum_{j=1}^{N_d} d_j^{\frac{1}{T}}} \quad (2)$$

where $p(f = f_i)$ represents the probability of selecting the i -th data. T denotes a balancing hyperparameter. N_d denotes the number of labeled data in \mathcal{D} . After each sampling step, the set \mathcal{D} and the distances are updated before selecting next data. This is iterated until the target number of data is selected.

The connection between the k-means++ initialization and the k-Center-Greedy algorithm becomes clear when considering different values of T . For $T \rightarrow 0$, $p(f)$ converges to a “one-hot” distribution, which precisely corresponds to the k-Center-Greedy algorithm. It greedily selects the sample whose embedding has the largest distance to the closest selected sample embedding. Similarly, the k-means++ initialization algorithm is equivalent to $T = 0.5$. It selects samples with probability proportional to the squared distance to the nearest selected sample embedding. In general, the hyperparameter T can also be thought of controlling the extent to which the diversity-based sampling process is randomized.

IV. DIVERSITY-BASED SEMI-SUPERVISED ACTIVE LEARNING

We introduce two important applications of diversity-based sampling for semi-supervised active learning. Firstly, we show that diversity-based sampling can be used for initial dataset selection on self-supervised representation. Secondly, a new semi-supervised active learning query strategy, which is based on diversity-based sampling from consistency-based embedding space is introduced.

A. DIVERSITY-BASED INITIAL DATASET SELECTION

Semi-supervised active learning methods generally start from a small labeled initial dataset. The methods iteratively select data over multiple active learning cycles for annotation. We propose a method that can select informative initial datasets for improving the whole active learning steps. Figure 1 shows an overview of the proposed initial dataset selection algorithm pipeline.

Basically, the diversity-based selection can be designed based on the features extracted from the trained model in the current active learning stage. Contrary to query strategies in active learning, there is no trained model at the initial dataset selection stage. Therefore, it is challenging to assess the diversity of data without having access to any label at all. The advances in self-supervised learning [11], [16], [18] have shown that it is feasible to learn meaningful representations from unlabeled data. It motivates our approach to initial dataset selection. The proposed method consists of representation learning step and diversity-based sampling step.

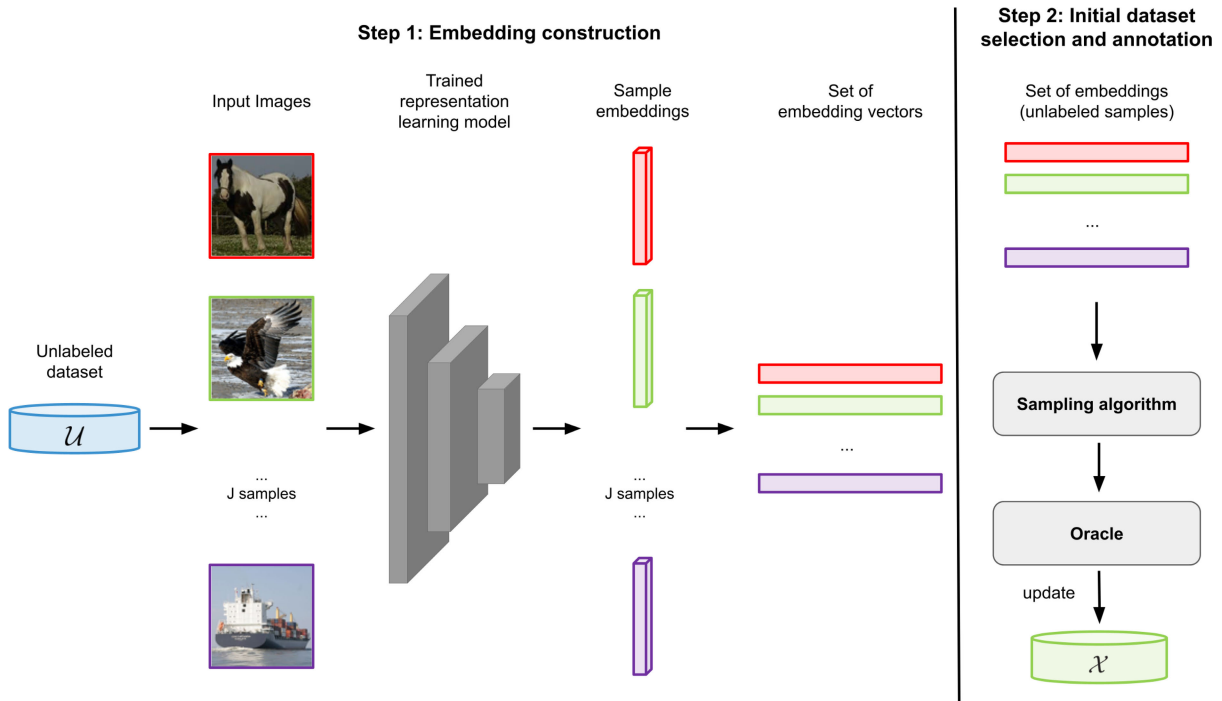


FIGURE 1. Overview of the initial dataset selection algorithm pipeline.

1) REPRESENTATION LEARNING

A self-supervised representation is used to calculate the feature vectors from unlabeled data. Depending on the characteristics of data, it is possible to choose proper representation.

2) DIVERSITY-BASED SAMPLING

In this step, the method selects an informative set of data based on the self-supervised representation. The diversity-based sampling is used to choose diverse informative data for training a model. In this study, we empirically use the hyperparameter T of 0.5 to select informative initial samples for labeling. These two steps constitute our diversity-based approach to the initial dataset selection for active learning. Section VI provides a thorough analysis of diversity-based sampling in this context.

B. DIVERSITY-BASED QUERY STRATEGY IN SEMI-SUPERVISED ACTIVE LEARNING

We further introduce the application in the query strategy of semi-supervised active learning. Reference [15] have proposed using *consistency*, referring to the consistency of model predictions on augmented input images, as the selection criterion. They assume that the unlabeled samples with highly inconsistent predictions are not useful for semi-supervised learning which is designed from the consistency regularization [6], [7], [37], [39]. In other words, if the model’s predictions are inconsistent after model training, it is reasonable to assume that the semi-supervised learning algorithm cannot extract useful information from that sample. Therefore, querying its label can be expected to be highly informative.

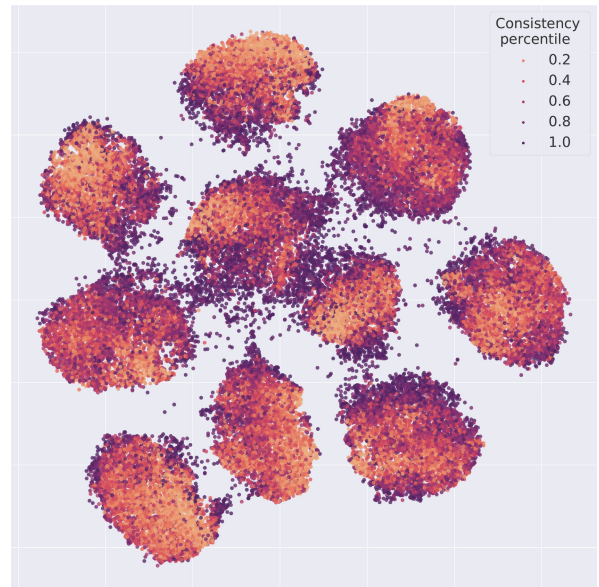


FIGURE 2. Heatmap of TSNE-embeddings [41] colored according to percentile of consistency scores (CIFAR-10).

However, [15] does not explicitly consider the diversity. As in [4] and [33], this might result in high overlaps of the data in a selected batch for annotation. Figure 2 illustrates the problem by showing a heatmap of TSNE-embeddings colored according to the consistency of model predictions. The embeddings and model predictions are based on an initial model trained using MixMatch with 150 labeled initial samples (randomly selected) on CIFAR-10. One can observe

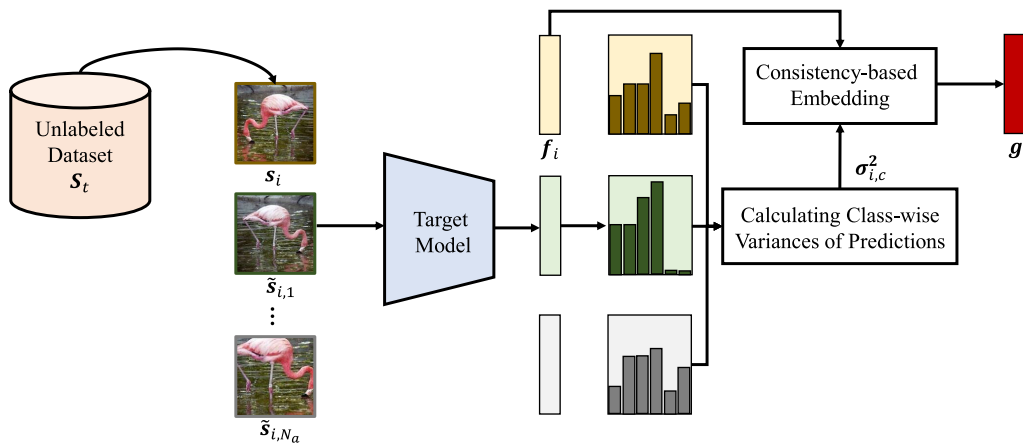


FIGURE 3. Overview of the proposed consistency-based embeddings query strategy for semi-supervised active learning.

that samples, on which model predictions are highly inconsistent are concentrated in small areas of embedding space. Interestingly, a key strength of recently proposed state-of-the-art supervised active learning algorithms such as BADGE [4] and SRAAL [45] lies in the effective combination of proven selection criteria such as diversity and uncertainty. Hence, combining diversity and consistency-based sample selection for semi-supervised active learning appears to be a highly promising approach. This critical insight inspires the semi-supervised active learning query strategy we introduce in the following. By applying diversity-based sampling on the consistency-based embedding, it is possible to consider both consistency and diversity for active learning. Let $\mathcal{S}_t = \{s_i : i \in (1, \dots, N_t)\}$ denote the pool of unlabeled samples at active learning step t . Let M_t denote the trained target model at active learning step t . The consistency of model M_t on an unlabeled sample s_i can be calculated by class-wise variances of predictions $\sigma_{i,c}^2$ on N_a augmented versions of input images $\tilde{s}_{i,k} = T(s_i)$ as:

$$\sigma_{i,c}^2 = V [P_{M_t}(y_c|s_i), P_{M_t}(y_c|\tilde{s}_{i,1}), \dots, P_{M_t}(y_c|\tilde{s}_{i,N_a})] \quad (3)$$

$T(\cdot)$ denotes a standard transformation operation for data augmentation. $V[\cdot]$ is the function for calculating variance and $\sigma_i^2 = (\sigma_{i,1}^2, \dots, \sigma_{i,N_c}^2)$ is the vector where the c -th element is defined as the variance of model predictions for class $c \in \{1, \dots, N_c\}$. N_c denotes the number of class for the dataset.

Given a sample s_i , the activations of the penultimate layer of the target model, denoted by f_i , encode sample-specific representation [33]. We can define a consistency-embedding g_i for data s_i as the last-layer activation f_i scaled by the sum of class-wise variances of predictions as

$$g_i = \left(\sum_{c=1}^{N_c} \sigma_{i,c}^2 \right) \cdot f_i \quad (4)$$

$$\|g_i\|_2 = \left(\sum_{c=1}^{N_c} \sigma_{i,c}^2 \right) \cdot \|f_i\|_2 \quad (5)$$

The size of consistency-based embedding is same with the dimension of the last-layer feature vector. By construction, the norm of the consistency-based embeddings is proportional to the sum of class-wise prediction variances. The diversity-based sampling algorithm is used on consistency-based embedding to select samples. These procedures are also illustrated in Figure 3.

As in [4], diversity-based algorithms select diverse and high-magnitude embeddings. In other words, the sampling algorithm is encouraged to select both diverse and inconsistent data where the model's predictions are inconsistent and having large embedding norms.

V. EXPERIMENTAL CONDITIONS

A. SETTINGS AND IMPLEMENTATION

1) DATASETS

In this study, experiments are conducted on two image classification datasets, CIFAR-10 [24] and Caltech-101 [13]. The CIFAR-10 dataset consists of 60,000 images of size 32×32 with 10 classes. The images are split into a train set of 50,000 images and a test set of the remaining 10,000 images. The Caltech-101 dataset consists of 8,677 images with 101 classes. The size of each image is roughly 300×200 pixels. The images are resized to 224×224 . The dataset is randomly split into 90% of training set and 10% of test set.

2) IMPLEMENTATION

Following previous work on semi-supervised active learning [15], [38], MixMatch [7] is used for model training. A Wide ResNet-28-2 [44] is used as network architecture for CIFAR-10 experiments. The default MixMatch hyperparameters for CIFAR-10 are kept for semi-supervised model training. A ResNet-18 is used for Caltech-101 experiments. The MixMatch hyperparameters used on Caltech-101 are adopted from default setting originally presented for CIFAR-100 [24], a dataset with a comparable number of classes. All considered algorithms, i.e. also supervised active learning algorithms, are run with the same learning rate of 0.002 and weight decay

of 0.00004 (default MixMatch settings [7]) using Adam optimizer [23]. The size of the initial-labeled dataset and the budget sizes, i.e. the number of samples selected for labeling at every active learning cycle, were chosen as in [15]. For Caltech-101 dataset, the initial dataset size was 388 (5%) and the budget size was 5% in each active learning step. For CIFAR-10 dataset, the initial dataset size was 150 (0.3%) and the budget sizes were [50 (0.1%), 50 (0.1%), 250 (0.5%), 250 (0.5%), 250 (0.5%)]. Due to the reason that model achieved high accuracy with a small number of labeled samples on CIFAR-10, we set the active learning configuration in this way.

The standard augmentation operation with random horizontal flips and random crops is used for the calculation of class-wise prediction variances (please see Equation 3). For calculating class-wise variances, the model predictions are obtained from 50 random augmentations on CIFAR-10 dataset [15] and 10 random augmentations on Caltech-101 dataset.

3) BASELINE ALGORITHMS

For comparison, we consider four different baseline query strategies. Maximum entropy [21] is a purely uncertainty-based baseline method, which selects data with the highest entropy at every active learning cycle. *Coreset* [33] selects the data based on the activations of the penultimate network layer as embedding and K-center-Greedy algorithm for considering diversity. *BADGE* [4] constructs gradient embedding to encode both diversity and uncertainty. The expected model change is used for calculating uncertainty. The k-means++ initialization algorithm is applied for balancing diversity and uncertainty on the gradient embedding. The supervised learning is used for Maximum entropy, Coreset and BADGE for model training. The training starts from random initial dataset. The data augmentation is conducted on each image by horizontally flipping and random cropping for supervised learning. Consistency [15] uses the sum of class-wise variance of the predictions over augmented versions of input image as the selection criterion. MixMatch [7] is adopted as for semi-supervised learning and the training is started from random initial dataset. We follow public implementations [3], [34].

4) EVALUATION

The empirical validation and qualitative analysis of the method proposed in this work is conducted based on a set of metrics introduced in the following.

In addition to comparing algorithms solely based on test accuracy, we analyze active learning query strategies with respect to characteristic features such as *consistency* and *diversity* following [15]. Unless otherwise stated, these metrics are computed as:

- **Consistency:** The consistency of model predictions on a given sample is computed based on predictions on augmented versions of the image. More formally, the

consistency of model predictions of sample u_i , which is denoted by c_i , is defined as the sum of class-wise prediction variances computed over predictions on multiple augmented versions of that sample, i.e. $c_i = \sum_{k=1}^K \sigma_{i,k}^2$ (see Equation 3). Low values of c_i indicate very consistent model predictions, while higher values of c_i indicate inconsistent model predictions on the unlabeled sample u_i .

- **Diversity:** The diversity of a batch of samples is defined as the average, pairwise L_2 -distance between sample embeddings, which are given by activations of the penultimate network layer. Low values of the diversity measure indicate a potentially high overlap in representational information of the samples in the batch. By contrast, high values of the diversity measure suggest a low overlap in representational information and have shown to be indicative of highly informative batches.

Based on the general experimental setting described above, the following sections present empirical evidence validating all design choices as well as showing the effectiveness of the proposed approach.

VI. EXPERIMENTAL RESULTS

A. DIVERSITY-BASED SAMPLING FOR INITIAL DATASET SELECTION

In the following, we provide an empirical analysis of the diversity-based approach to initial dataset selection. The analysis is conducted on CIFAR-10 and RotNet [16] is used to generate sample embeddings. The default network architecture as well as hyperparameters are kept for training. In accordance with [16], the activations after the second layer of the network are used to compute sample embeddings. As the second layer activations are of dimension $192 \times 8 \times 8$, a simple flattening operation would yield embedding vectors of dimension 12288×1 . As discussed in [1], distance metrics such as the euclidean distance behave oddly in very high-dimensional spaces and therefore are not suitable as input to a diversity-based sampling algorithm. Hence, a global average pooling operation is applied to the second layer activations yielding sample embeddings of dimension 192×1 . This addresses the curse of dimensionality and reduces the computational time required by the sampling step. We evaluate the effectiveness of the diversity-based initial dataset selection algorithm for different values of T (see Equation 2) based on two critical metrics. Firstly, we evaluate the sample diversity of selected initial datasets, which is defined as the average L_2 -distance between RotNet-embeddings of selected samples [15]. Secondly, we evaluate to what extent the class distribution of initial datasets selected by our approach matches the class distribution of the source dataset, i.e. CIFAR-10. For this purpose, we introduce a distance measure d , referred to as *class-distribution distance* in the following. Let $\mathcal{X} = \{(x_i, y_i) : i \in (1, \dots, N)\}$ denote a selected initial dataset with samples x_i and corresponding labels y_i . Then the

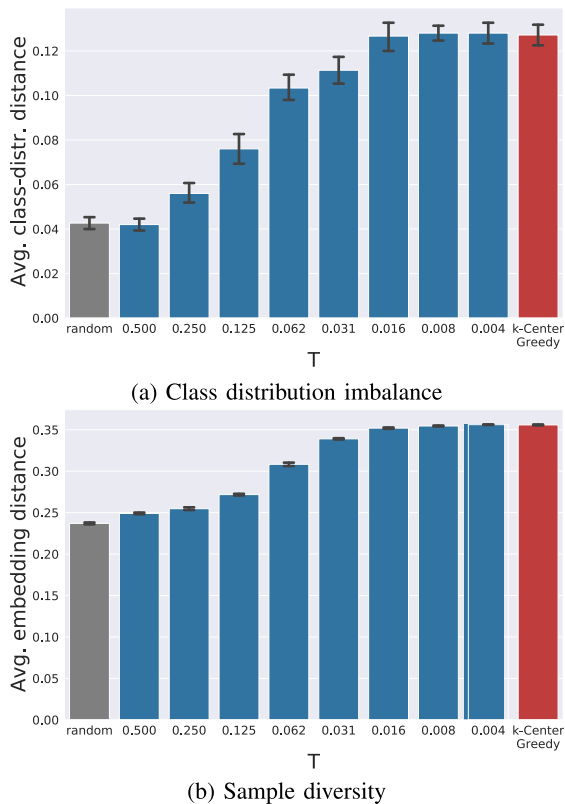


FIGURE 4. Analysis of our diversity-based initial dataset selection algorithm with respect to (a) class imbalance and (b) diversity for different values of T . The bars represent the standard error.

class-distribution distance is defined as

$$d = \max_c |\tilde{p}_c - p_c|, \tag{6}$$

where $\tilde{p}_c = \frac{1}{|\mathcal{X}|} \sum_{i=1}^{|\mathcal{X}|} \mathbb{1}\{y_{i,c} = 1\}$, $y_{i,c}$ is the c -th element of the one-hot label of sample x_i and p_c is the true proportion of samples belonging to class c in the source dataset (e.g., p_c is equal to 10% for all c in CIFAR-10).

Figure 4 presents results on the selection of 150 initial samples on CIFAR-10 and averaged over ten runs for each sampling algorithm. It compares random sampling, the k -Center-Greedy algorithm ($T \rightarrow 0$) and the formulation of the diversity-based sampling algorithm with respect to the previously introduced metrics. Note that the k -means++ initialization step is equivalent to $T = 0.5$. k -means++ initialization step selects initial datasets, whose class distribution is as close to the class distribution of the source dataset as the class distribution of randomly selected datasets. At the same time, it succeeds at sampling a more diverse set of initial samples than random selection. The difference is statistically significant ($p < 0.001$ by t-test). The k -Center-Greedy algorithm, by contrast, selects highly imbalanced initial datasets and is therefore not suitable for the initial dataset selection.

For qualitative analysis of self-supervised representation, we visualize the feature embeddings. Figure 5 shows TSNE-embeddings of the representations learned

by the self-supervised representation learning algorithms. For CIFAR-10 dataset, RotNet [16] is used for calculating features. For Caltech-101 dataset, BYOL [18] which is pre-trained from ImageNet dataset is used. As shown in the figure, for both datasets, feature representations from the same classes are clustered and different classes are discriminated to some degree although the feature embedding model is trained without any labels in the target dataset (for CIFAR-10) or only pre-trained from ImageNet (for Caltech-101).

B. DIVERSITY-BASED SAMPLING FOR CONSISTENCY-BASED EMBEDDINGS QUERY STRATEGY

This section provides empirical analysis of the semi-supervised active learning query strategy introduced in Subsection IV-B, which applies the diversity-based sampling algorithm to the consistency-based embeddings. In particular, we analyze samples selected by the query strategy with respect to consistency and diversity for different values of T . Furthermore, we show that the consistency-based embeddings query strategy succeeds at balancing both of these selection criteria when compared to baseline active learning algorithms.

As in Subsection VI-A, diversity is computed as the average L_2 -distance between the sample embeddings, which, in this context, are given by the activations of the penultimate network layer of the trained model (at the current active learning cycle). The consistency of model predictions is computed as the sum of class-wise prediction variances on augmented versions of a given sample (see Equation 3). The following analysis is conducted on CIFAR-10 and all considered query strategies start from the same initial models trained on a randomly selected initial labeled dataset using MixMatch. In accordance with the general experimental setting on CIFAR-10, the initial dataset is of size 150 and all query strategies are run with a budget size of 50. We compute the metrics based on samples selected by each query strategy and averaged over five independent runs.

1) ANALYSIS OF DIVERSITY-BASED SAMPLING FOR OUR QUERY STRATEGY

We start by evaluating our query strategy, which uses consistency-based embeddings and the diversity-based sampling algorithm, for different values of T . Figure 6 shows that using lower values for T leads to selecting more diverse samples as well as samples, on which model predictions are more inconsistent. On the basis of these results, the k -Center-Greedy algorithm is identified as the preferred sampling algorithm in the context of the consistency-based embeddings query strategy and used to obtain the experimental results presented in Subsection VI-C.

2) BALANCING SELECTION CRITERIA

The critical weakness of purely consistency-based query strategy [15] is, that it does not explicitly consider the diversity of representational information for sample selection. Therefore, it is interesting to study to what extent



(a) Class-wise TSNE-embeddings based on RotNet [17] (CIFAR-10)

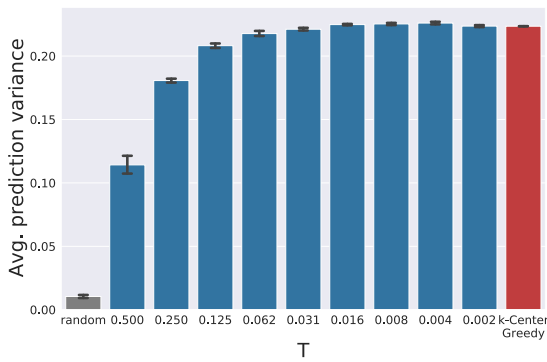


(b) Class-wise TSNE-embeddings based on BYOL [19] (Caltech-101)

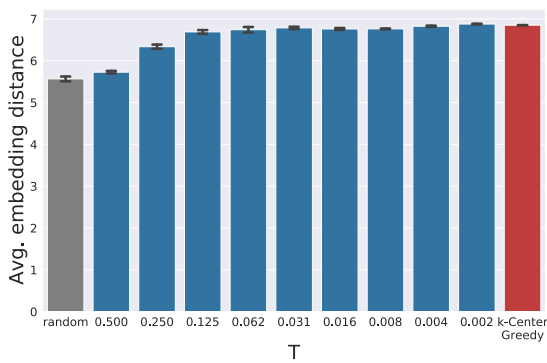
FIGURE 5. Qualitative analysis of self-supervised representation learning algorithms. The representations learned by self-supervised learning on (a) CIFAR-10 and (b) Caltech-101 are qualitatively analyzed based on their TSNE-embeddings. The representations on CIFAR-10 are generated using RotNet, while the representations on Caltech-101 are generated using a BYOL model, which is pretrained on ImageNet.

the proposed query strategy with consistency-based embeddings can address this issue. The core idea of the proposed query strategy with consistency-based embeddings is to naturally combine effective selection criteria by applying diversity-based sampling to a consistency-based embedding space. We demonstrate the effectiveness of this approach based on a comparison with the query strategies of baseline algorithms, namely Maximum Entropy [21], Coreset [33], BADGE [4] and consistency-based semi-supervised active learning [15]. Note that in this context, we consider the characteristics of samples selected by the *query strategies* of the baseline algorithms in an initial sampling step. Figure 7a shows the comparison of baseline algorithms with respect to the *consistency* of model predictions of samples they select. By construction, the purely consistency-based query strategy [15] (referred to as “Consistency” in the figure) selects

the samples with the highest prediction variance. Contrary to that, active learning algorithms originally developed for the supervised active learning setting tend to select samples with significantly lower prediction variance. Furthermore, one can observe that, as intended, the consistency-based embeddings query strategy (referred to as “Ours”) selects samples with the second highest prediction variances on average. Figure 7b illustrates a comparison of all selected baseline algorithms with respect to sample *diversity*. *Coreset* selects batches for having the highest sample diversity by construction, while batches selected by other algorithms exhibit significantly lower sample diversity. Our query strategy selects samples with the second highest diversity score. Overall, the empirical analysis highlights that, as intended, our query strategy succeeds at balancing diversity and consistency as selection criteria for semi-supervised active learning.



(a) Consistency



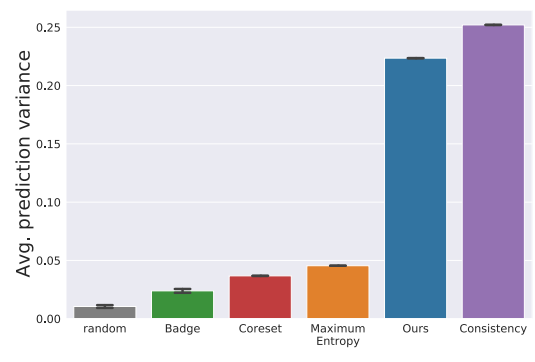
(b) Sample diversity

FIGURE 6. Empirical analysis of the consistency-based embeddings query strategy with respect to (a) consistency and (b) diversity of selected samples for different values of T .

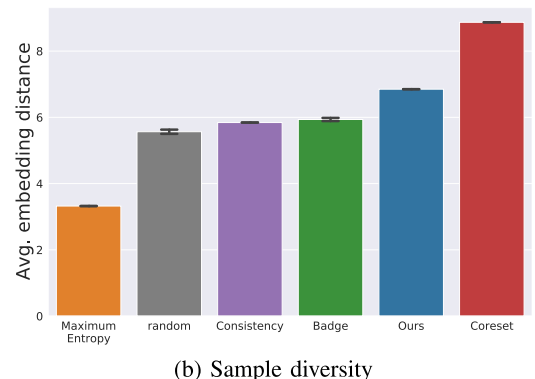
C. EVALUATION ON ACTIVE LEARNING

We compare the active learning performance compared with other active learning approaches including maximum entropy [21], Coreset [33] and BADGE [4] as well as consistency-based semi-supervised active learning [15]. Please note that maximum entropy, Coreset and BADGE rely on supervised learning. Consistency-based semi-supervised active learning and our algorithm use semi-supervised learning (i.e., MixMatch [7] in this study). Active learning steps start from initial models trained on the same initial datasets. Random selection is used for setting initial datasets. For the case of BADGE, the computational complexity increases with respect to the number of classes, which makes it difficult to apply BADGE to Caltech-101 setting. As a result, we only conduct BADGE experiments on CIFAR-10. We follow the testing protocol in [6], [7], and [37] where the test accuracy is calculated from an exponential moving average of model parameters. We use the median accuracy of the last 20 epochs on CIFAR-10 [7] and the median accuracy of the last 6 epochs on Caltech-101. We repeat the experiments over five times by changing the random seeds and calculate the average accuracy from five runs.

Note that our initial dataset selection algorithm uses embeddings which are obtained from RotNet [16] for CIFAR-10 setting. For Caltech-101 setting, which is a more challenging dataset, we use *BYOL* [18] with a ResNet-50



(a) Consistency

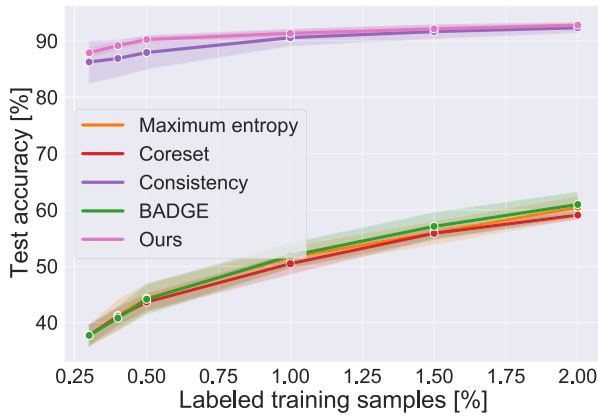


(b) Sample diversity

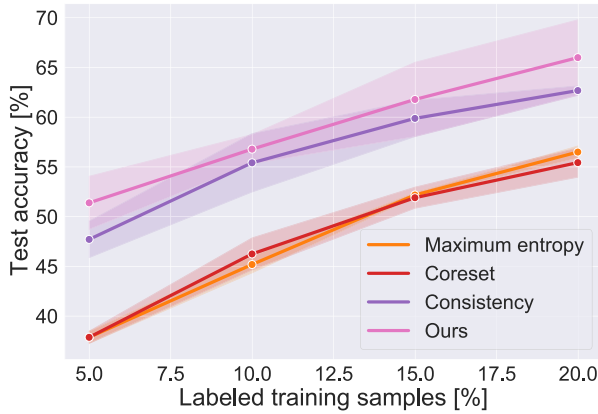
FIGURE 7. Comparison of active learning query strategies based on diversity and consistency.

pretrained on Imagnet [42] for obtaining embeddings from samples. The activations of the penultimate network layer (i.e., dimension of 2024) are used as sample representations.

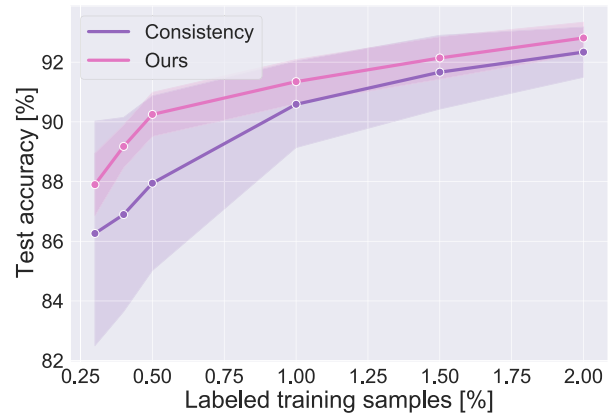
Figure 8 shows the results of comparison. As shown in the figure, our approach outperforms other active learning methods over multiple active learning stpes on both CIFAR-10 and Caltech-101 datasets. Semi-supervised active learning approaches outperform supervised active learning algorithms. Figure 9 shows the detail comparison of our approach with consistency-based approach [15]. As shown in the figure, On CIFAR-10, the proposed method achieves an accuracy of 92.81% by only labeling 2% of training data. The consistency-based approach [15] achieves 91.66% with the same annotation cost. On Caltech-101, our method can achieve an accuracy of 65.99% by only labeling 20% of training data which is higher than the accuracy of consistency-based approach [15] (i.e., 62.68% with the same annotation cost). Please note that supervised active learning algorithms achieve accuracies of approximately 60% on CIFAR-10 and 56% on Caltech-101 with the same annotation cost. These results show the efficiency of our method in terms of annotation cost. The major difference between our method and consistency-based approach [15] is that our method can consider the diversity of samples during the selection. As shown in Figure 7(b), our method can balance consistency and diversity during the sample selection by using diversity sampling on consistency-based embeddings.



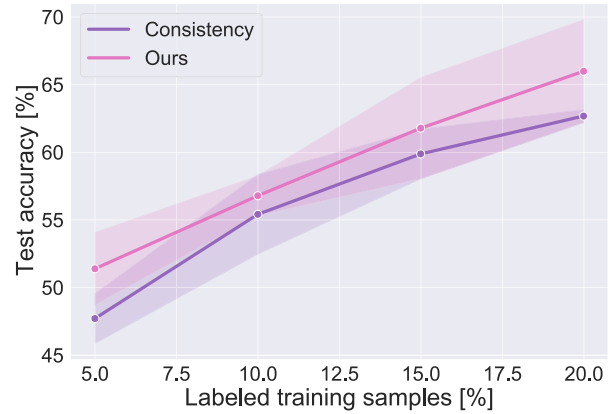
(a) CIFAR-10



(b) Caltech-101



(a) CIFAR-10



(b) Caltech-101

FIGURE 8. Comparison of the proposed approach to baseline algorithms based on active learning performance.

D. ABLATION STUDY

In this subsection, we provide further empirical evidence validating the effectiveness of the initial dataset selection algorithm and the consistency-based embeddings query strategy, separately.

1) EFFECTIVENESS OF THE INITIAL DATASET SELECTION ALGORITHM

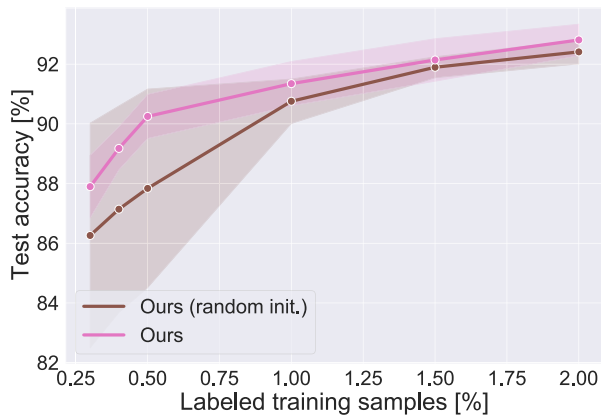
Figure 10 shows a comparison of the performance obtained in the two different settings on both CIFAR-10 and Caltech-101. As shown in the figure, the proposed initial sample selection strategy improves the accuracy of initial models. The accuracy of initial model from our approach is 87.90% on CIFAR-10 which is higher than random initialization where the accuracy is 86.26%. Also, the accuracy is 51.39% with our method which is higher than random initialization which achieves the accuracy of 47.71% on Caltech-101. These improvements results in higher accuracies over following active learning steps. This is mainly due to the reason that Our approach can select more diverse sample compared to random selection (please see Figure 4). In the same time, the class distribution imbalance is comparable between our approach and random selection which is also important to have a good starting point for active learning.

FIGURE 9. Comparison of the proposed approach to pure consistency-based approach [15] on the active learning performance.

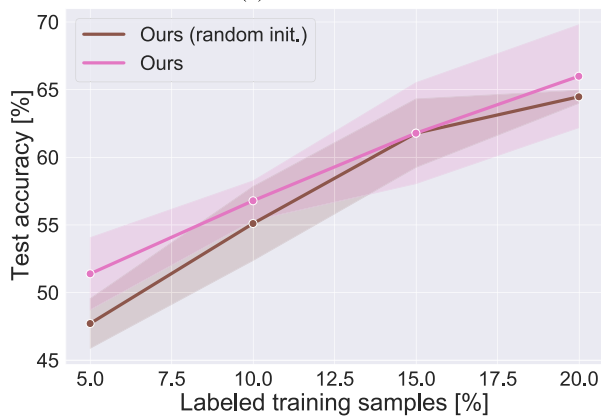
2) EFFECTIVENESS OF THE CONSISTENCY-BASED EMBEDDINGS QUERY STRATEGY

Figure 11 shows the performance of the proposed consistency-based embeddings semi-supervised active learning algorithm compared to the consistency-based semi-supervised active learning algorithm on CIFAR-10 and Caltech-101. It is important to note, that both approaches start from the exact same initial models trained on randomly selected initial datasets in order to guarantee comparability. In the context of this comparison, the approach proposed in this work does therefore not employ the diversity-based initial dataset selection algorithm.

On CIFAR-10, the proposed algorithm achieves an accuracy of 92.41% by only using 2% of labeled data, i.e. the equivalent of 100 labeled images per class. For reference, the purely consistency-based semi-supervised active learning algorithm [15] achieves an accuracy of 92.33%. Furthermore, the proposed algorithm slightly outperforms purely consistency-based semi-supervised active learning on 4 of 5 active learning steps. The largest difference between the two algorithms can be observed on the fourth active learning step: the proposed algorithm achieves an accuracy of 91.90% compared to an accuracy of 91.66% achieved by the purely consistency-based algorithm. Similarly, the proposed



(a) CIFAR-10



(b) Caltech-101

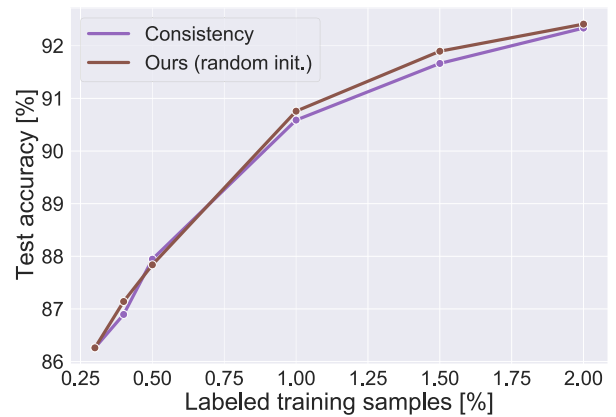
FIGURE 10. Validation of the initial datasets selection algorithm based on test accuracy.

algorithm outperforms purely consistency-based semi-supervised active learning on 2 out of 3 active learning steps on Caltech-101. On the last step, it reaches an accuracy of 64.47% with around 20% of labeled data, i.e. the equivalent of roughly 15 images per class on average. In comparison, the purely consistency-based algorithm achieves an accuracy of 62.68% on the final step.

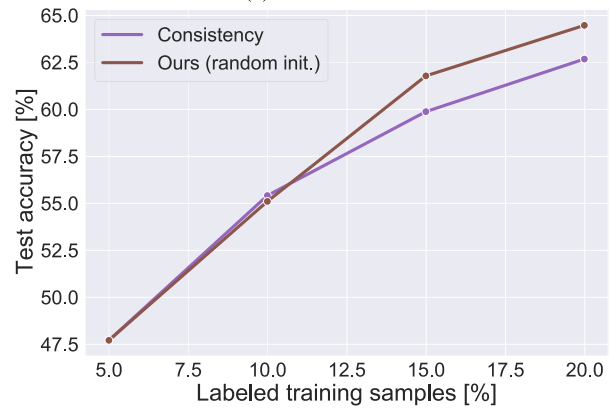
VII. DISCUSSION

Experimental results of this study verify the effectiveness of the proposed diversity-based sampling in semi-supervised active learning. For the applications, we show two interesting directions of initial dataset selection and a new query strategy for semi-supervised active learning.

Interestingly, we found that introducing randomization, i.e. using the k-means++ initialization step for sample selection, increased the effectiveness of the initial dataset selection algorithm (see Subsection VI-A). It is reasonable to assume that the extent to which randomness is beneficial in this context strongly depends on the quality of sample embeddings. More specifically, we expect that the better the learned representations capture semantic information the less randomization is required to ensure the robustness and effectiveness



(a) CIFAR-10



(b) Caltech-101

FIGURE 11. Validation of the consistency-based embeddings query strategy based on test accuracy.

of the initial dataset selection algorithm. Future work might further explore the influence of the quality of learned sample embeddings on the effectiveness of the proposed initial dataset selection algorithm.

Our choice of the initial dataset size was guided by previous works [7], [15]. However, ideally, active learning algorithms should be started with the smallest possible initial dataset size in order to optimally use the learning-based sample selection in subsequent cycles. At the same time the initial dataset has to be large enough to avoid the cold-start problem discussed in [15], i.e. large enough to ensure the convergence of semi-supervised learning. Gao et al. [15] provide an exploratory analysis of optimal start sizes based on random initial dataset selection. We believe that future research using our representation-based approach to initial dataset selection can play a critical role in further improving the label-efficiency of deep learning.

We acknowledge that the presented analysis and results are limited to the task of image classification on well-known benchmark datasets. Therefore, an interesting direction for future research would be generalizing the presented concepts to alternative datasets and tasks such as semantic segmentation.

VIII. CONCLUSION

We introduced diversity-based sampling algorithms for two important steps of semi-supervised active learning. The empirical analysis showed that introducing randomization to our diversity-based initial dataset selection algorithm increases its robustness as well as ensuring that selected samples encode diverse representational information and are balanced among classes. We proposed the consistency-based embeddings query strategy highlighting that diversity-based sampling can be applied to a specifically constructed embedding space in order to naturally balance effective selection criteria. Furthermore, we demonstrated that the empirical analysis of both proposed components translates into significant gains in performance in semi-supervised active learning for image classification. We hope the presented concepts inspire future research and are built upon to improve the label-efficiency of neural network training.

REFERENCES

- [1] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional space," in *Proc. Int. Conf. Database Theory*, 2001, pp. 420–434.
- [2] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2007, pp. 1027–1035.
- [3] J. T. Ash, C. Zhang, A. Krishnamurthy, J. Langford, and A. Agarwal. (2020). *Batch Active Learning by Diverse Gradient Embeddings*. [Online]. Available: <https://github.com/JordanAsh/badge>
- [4] J. T. Ash, C. Zhang, A. Krishnamurthy, J. Langford, and A. Agarwal, "Deep batch active learning by diverse, uncertain gradient lower bounds," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2020, pp. 1–26.
- [5] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
- [6] D. Berthelot, N. Carlini, E. D. Cubuk, A. Kurakin, K. Sohn, H. Zhang, and C. Raffel, "RemixMatch: Semi-supervised learning with distribution matching and augmentation anchoring," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2019, pp. 1–13.
- [7] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, "MixMatch: A holistic approach to semi-supervised learning," in *Proc. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 32, 2019, pp. 1–11.
- [8] Z. Borsos, M. Tagliasacchi, and A. Krause, "Semi-supervised batch active learning via bilevel optimization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 3495–3499.
- [9] F. Buchert, N. Navab, and S. T. Kim, "Exploiting diversity of unlabeled data for label-efficient semi-supervised active learning," in *Proc. 26th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2022, pp. 2063–2069.
- [10] O. Chapelle, B. Scholkopf, and A. Zien, "Semi-supervised learning," *IEEE Trans. Neural Netw.*, vol. 20, no. 3, p. 542, Mar. 2009.
- [11] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.
- [12] T. Drugman, J. Pylkkonen, and R. Kneser, "Active and semi-supervised learning in ASR: Benefits on the acoustic and language models," 2019, [arXiv:1903.02852](https://arxiv.org/abs/1903.02852).
- [13] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop*, 2004, p. 178.
- [14] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1050–1059.
- [15] M. Gao, Z. Zhang, G. Yu, S. O. Arik, L. S. Davis, and T. Pfister, "Consistency-based semi-supervised active learning: Towards minimizing labeling cost," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 510–526.
- [16] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–16.
- [17] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
- [18] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. A. Pires, Z. Guo, M. G. Azar, B. Piot, K. Kavukcuoglu, R. Munos, and M. Valko, "Bootstrap your own latent—A new approach to self-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 21271–21284.
- [19] Z. Hao, C. Lu, Z. Huang, H. Wang, Z. Hu, Q. Liu, E. Chen, and C. Lee, "ASGN: An active semi-supervised graph neural network for molecular property prediction," in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 731–752.
- [20] G. He, B. Li, H. Wang, and W. Jiang, "Cost-effective active semi-supervised learning on multivariate time series data with crowds," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 52, no. 3, pp. 1437–1450, Mar. 2022.
- [21] A. J. Joshi, F. Porikli, and N. Papanikolopoulos, "Multi-class active learning for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2372–2379.
- [22] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3128–3137.
- [23] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–15.
- [24] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Univ. Toronto, Toronto, ON, Canada, Tech. Rep. TR-2009, 2009.
- [25] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Int. Conf. Mach. Learn. Workshop*, 2013, vol. 3, no. 2, p. 896.
- [26] Y. Li, J. Yin, and L. Chen, "SEAL: Semisupervised adversarial active learning on attributed graphs," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 3136–3147, Jul. 2021.
- [27] K. Madhawa and T. Murata, "Metal: Active semi-supervised learning on graphs via meta-learning," in *Proc. Asian Conf. Mach. Learn.*, 2020, pp. 561–576.
- [28] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2536–2544.
- [29] P. K. Rhee, E. Erdenee, S. D. Kyun, M. U. Ahmed, and S. Jin, "Active and semi-supervised learning for object detection with imperfect data," *Cogn. Syst. Res.*, vol. 45, pp. 109–123, Oct. 2017.
- [30] D. Roth and K. Small, "Margin-based active learning for structured output spaces," in *Proc. Eur. Conf. Mach. Learn.*, 2006, pp. 413–424.
- [31] M. Sajjadi, M. Javanmardi, and T. Tasdizen, "Regularization with stochastic transformations and perturbations for deep semi-supervised learning," in *Proc. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 29, 2016, pp. 1–9.
- [32] T. Scheffer, C. Decomain, and S. Wrobel, "Active hidden Markov models for information extraction," in *Proc. Int. Symp. Intell. Data Anal.*, 2001, pp. 309–318.
- [33] O. Sener and S. Savarese, "Active learning for convolutional neural networks: A core-set approach," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–13.
- [34] O. Sener and S. Savarese. (2020). *Active Learning Via Core-Sets*. [Online]. Available: https://github.com/ozansener/active_learning_coreset
- [35] B. Settles, "Active learning literature survey," Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep. 1648, 2009.
- [36] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul. 1948.
- [37] K. Sohn, D. Berthelot, C.-L. Li, Z. Zhang, N. Carlini, E. D. Cubuk, A. Kurakin, H. Zhang, and C. Raffel, "FixMatch: Simplifying semi-supervised learning with consistency and confidence," in *Proc. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 33, 2020, pp. 596–608.
- [38] S. Song, D. Berthelot, and A. Rostamizadeh, "Combining MixMatch and active learning for better accuracy with fewer labels," 2019, [arXiv:1912.00594](https://arxiv.org/abs/1912.00594).
- [39] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 30, 2017, pp. 1–10.

[40] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *J. Mach. Learn. Res.*, vol. 2, pp. 45–66, Nov. 2001.

[41] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–27, 2008.

[42] X. Yao. (2020). *Byol-PyTorch*. [Online]. Available: <https://github.com/yaox12/BYOL-PyTorch>

[43] D. Yoo and I. S. Kweon, "Learning loss for active learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 93–102.

[44] S. Zagoruyko and N. Komodakis, "Wide residual networks," in *Proc. Brit. Mach. Vis. Conf.*, 2016, pp. 1–15.

[45] B. Zhang, L. Li, S. Yang, S. Wang, Z.-J. Zha, and Q. Huang, "State-relabeling adversarial active learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8756–8765.

[46] H. Zhang, M. Cisse, N. Y. Dauphin, and D. Lopez-Paz, "Mixup: Beyond empirical risk minimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2018, pp. 1–13.



NASSIR NAVAB (Fellow, IEEE) received the Ph.D. degree from INRIA, University of Paris XI, France. He has secondary faculty appointments at both affiliated Medical Schools. In 1994, he joined at Siemens Corporate Research (SCR), where he was a Distinguished Member. He is currently a Full Professor and the Director of the Laboratory for Computer-Aided Medical Procedures with Johns Hopkins University, and the Technical University of Munich. He is the author of 100 peer-reviewed scientific articles, with more than 58,000 citations, and an H-index of 105 as of December 2022. He is the author of more than 30 awarded papers including 11 at MICCAI, five at IPCAI, and three at IEEE ISMAR. He is the Inventor of 50 granted U.S. patents and more than 50 international ones. His current research interests include medical augmented reality, computer-aided surgery, medical robotics, and machine learning. In 2012, he was elected as a fellow of the MICCAI Society. He has acted as a member of the Board of Directors of the MICCAI Society, from 2007 to 2012 and from 2014 to 2017, and serves on the Steering Committee of the IEEE Symposium on Mixed and Augmented Reality (ISMAR) and Information Processing in Computer-Assisted Interventions (IPCAI). He received two years of a Postdoctoral Fellowship from the MIT Media Laboratory and the Siemens Inventor of the Year Award, in 2001. He received the SMIT Society Technology Award for the introduction of Camera Augmented Mobile C-arm and Freehand SPECT Technologies, in 2010, and the 10 Years Lasting Impact Award from IEEE ISMAR, in 2015.



FELIX BUCHERT received the B.S. degree in engineering science and the M.S. degree in data engineering and analytics from the Technical University Munich, Munich, Germany. He was a Visiting Student at Université Laval, QC, Canada. His current research interests include the broad area of machine learning and deep learning with particular focus on active learning, semi-supervised learning, representation learning, and self-supervised learning.



SEONG TAE KIM (Member, IEEE) received the B.S. degree from Korea University, Seoul, South Korea, and the M.S. and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2014 and 2019, respectively. In 2015, he was a Visiting Researcher at the University of Toronto, ON, Canada. From 2019 to 2021, he was a Senior Research Scientist at the Chair for Computer Aided Medical Procedures, Technical University of Munich, Munich, Germany. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Kyung Hee University, Yongin-si, South Korea. He has authored or coauthored more than 50 peer-reviewed journals and conference papers. His current research interests include deep learning, spatio-temporal learning, explainable deep learning, and data-efficient deep learning. He received the Best Student Paper Award of SPIE Medical Imaging, in 2018.

• • •